Trop. Med., 34 (2), 39−54, June, 1992

Comparative Nucleotide and Amino Acid Sequences of the Envelope Protein of Five Hepatitis C Viral cDNA Clones Isolated from HCV Infected Patients in Nagasaki Prefecture

Shintaro KURIHARA

Department of Virology, Institute of Tropical Medicine, Nagasaki University, 12−4 Sakamoto-machi, Nagasaki 852, Japan

Abstract: cDNA fragments covering the whole envelope (E) protein region of hepatitis C virus (HCV) have been molecularly cloned by using polymerase chain reaction (PCR) technique from the serum of five HCV infected patients in Nagasaki Prefecture. Their nucleotide and encoded amino acid sequences were determined and compared with other HCV cDNA clones isolated in Japan and the United States. The result showed that HCV cDNA isolates compared in this study were roughly divided into two subgroups; Japanese subgroup and USA subgroup. All HCV cDNA isolates in Nagasaki Prefecture were included in the former subgroup. Amino acid homology analysis showed that conformation of the E protein was relatively conserved among all compared HCV cDNA isolates, indicating that the E protein may be under structural or functional constraint.

Key words: Hepatitis C virus, Envelope protein, Nucleotide and amino acid sequences, Polymerase chain reaction

INTRODUCTION

Blood-borne non-A, non-B hepatitis (NANBH), which is caused by viral infection, is a major form of post-transfusion hepatitis (Feinstone, 1975) and frequently develops into chronic hepatitis and malignant liver diseases such as cirrhosis and hepatocellular carcinoma (Dienstag, 1983; Kiyosawa *et al.*, 1984, Koretz *et al.*, 1980; Realdi *et al.*, 1982; Resnick *et al.*, 1983). Recently, cDNA clones of probable hepatitis C virus (HCV) have been obtained by immunoscreening of cDNA library derived from the plasma of a chimpanzee chronically infected with HCV in the United States (Choo *et al.*, 1989). Using these clones, they subsequently developed an immunoassay to detect specific anti-HCV antibodies (anti-C100) (Kuo *et al.*, 1989). Seroepidemiological studies based on the detection of anti-C100 indicated that HCV not only is the major cause of post transfusion NANBH but also may have a pathogenic role in many cases of chronic liver disease (Alter *et al.*, 1989; Kuo *et al.*, 1989; Miyamura *et al.*, 1990; Plageman, 1991; Saito *et al.*, 1990). More recently, several research groups in

Received for Publication April 4, 1992.
Contribution No. 2709 from the Institute of Tropical Medicine, Nagasaki University.

Japan have isolated cDNAs spanning partial or total length of HCV genome directly from human patients by using probes based on the sequence of the prototype isolates in the United States (Kato *et al.*, 1990; Okamoto *et al.*, 1990; Takamizawa *et al.*, 1991; Takeuchi *et al.*, 1990). From their analyses on primary structure of the viral genome, the HCV genome is considered to be a positive-stranded RNA molecule of about 10,000 nucleotides which encodes a large polyprotein precursor of about 3,000 amino acids. A search for sequence homology of the HCV genome indicated that this virus is related to flaviviruses and pestiviruses (Houghton *et al.*, 1988; Miller and Purcell, 1990). The gene products of HCV have not been identified, but HCV genome organization was speculated as shown in Figure 1 (Takamizawa *et al.*, 1991), by comparing deduced amino acid sequence of the long open reading frame (ORF), its hydrophobicity profile, and the location of potential glycosylation sites with those of pesti- and flavivirus genomes. The HCV viral genome encodes one large polyprotein that is subsequently cleaved into three structural proteins, the core (C) protein, the matrix (M) protein, the envelope (E) protein, and six nonstructural proteins, NS1, NS2, NS3, NS4a, NS4b, and NS5. It is known that the epitopes which are recognized by the neutralizing antibodies exist on the E protein region in the case of flaviviruses. Therefore, the E protein region would become the first target for vacccine development of HCV infection. In this study, by using polymerase chain reaction (PCR), a technique for DNA amplification *in vitro* (Saiki *et al.*, 1988), a cDNA fragment of HCV genome which encodes the total E protein region have been molecularly cloned from the serum samples of 5 patinets with acute and chronic HCV infection in Nagasaki Prefecture. Their nucleotide and deduced amino acid sequences were comparatively analyzed with those of the HCV cDNAs isolated so far in other parts of Japan and in the United States in order to analyze the molecular epidemiology of local strains. These informations may also be useful for the strategies to develop the HCV vaccine.

Materials and Methods

Preparation of RNA from human serum samples; Serum samples obtained from NANBH infected patients in Nagasaki Prefecture which were kindly provided by the Second Department of Internal Medicine of Nagasaki University Hospital were used. The serum samples were first screened the presence of anti-C100 by using Ortho Diagnostic enzyme-linked immunosorbent assay kit according to the manufacturer's instruction. In the cases of chronic NANBH infection, anti-C100 positive serum samples were selected and used for RNA extraction. In the cases of acute NANBH infection, the serum samples were used for RNA extraction regardless of the presence of anti-C100 because it is known that the detection of anti-C100 during acute phase of HCV infection (especially within one month since HCV infecton) is usually difficult (Alter *et al.*, 1989; Kuo *et al.*, 1989). To extract RNA, 100 μl serum was made up to 200 μl with sterile distilled water and mixed with an equal volume of buffer containing 0.2 M "tris"−HCl (pH 7.5), 25 mM ethylenediaminetetra-acetic acid (EDTA), 0.3 M NaCl, 2% sodium dodecyl sulfate (SDS), and 200 μg/ml proteinase K, and in-

cubated at 37℃ for 40 min. Proteins were removed by two extractions with phenol and once with chloroform. Twenty micrograms glycogen was added to the aqueous phase and the RNA was precipitated by addition of three volumes of ice-cold absolute ethanol. After storage at −70℃ for 1 hr, the RNA was pelleted in an Eppendolf centrifuge (30 min, 15,000×*g*, 4℃). The pellet was washed once in 70% ethanol, vacuum dried, and dissolved in 50 μl sterile distilled water. RNA solution was stored at −70℃.

Primers for PCR reaction: The oligonucleotide primers were synthesized with an Applied Biosystems DNA synthesizer (model 380B) and confirmed for purity by ion-exchange gel chromatography (Gen-pack; Waters). The sequences of primers were 5'-GTAATTTGG-GTAAGGTCATCGATAC-3'(sense) and 5'-TGCCAACTCCCATTGGTGTT-3'(antisense). This primer pair was expected to amplify 868 base pair DNA from putative region covering the whole E protein (nucleotide position: 707-1574) according to the sequences reported by Takamizawa *et al.* (1991) (Fig. 1).

Reverse-transcription (RT)-PCR: Five microliters of RNA solution was added to 95 μl of RT-PCR mixture containing 100 pmol each primers, 0.2 mM each deoxynucleoside triphosphate, 10 mM Tris (pH 8.9), 1.5 mM $MgCl_2$, 80 mM KCl, 0.5 mg of bovine serum albumin per ml, 0.1% sodium cholate, 0.1% Triton X-100, 10 U of reverse transcriptase (Life Science Inc.), and 2 U of Tth DNA polymerase, a thermostable DNA polymerase (Toyobo Co.). The reaction mixture was covered by 2 drops of mineral oil, and the tube was incubated for 10 min at 53℃ for RT. PCR amplification (92℃ for 60 sec, 53℃ for 90 sec, and 72℃ for 120 sec by thermal cycler; Iwaki Co.) was started immediately after the RT and repeated 35 times. PCR product was subjected to 3% agarose gel electrophoresis, and amplified DNA fragments were visualized by ethidium bromide staining.

Cloning and sequencing of HCV cDNA: The cDNA fragment amplified by RT-PCR was excised from the agarose gel, phosphorylated with T4 polynucleotide kinase (Nippon Gene Co.) and blunted with T4 DNA polymerase (Takara Co.). The modified DNA fragment was subcloned into the SmaI site of pUC19. The nucleotide sequence of the isolated clone was determined by the dideoxy chain termination method (Sanger *et al.*, 1977).

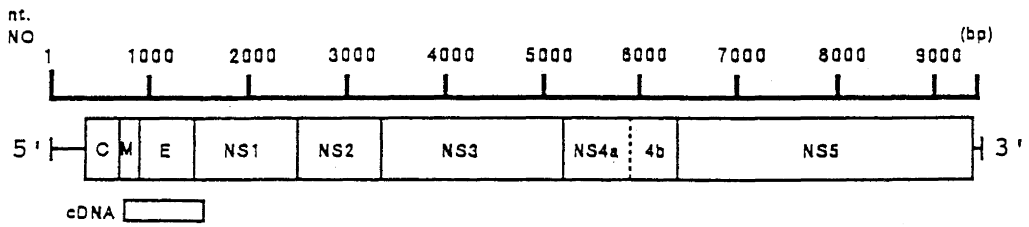

Fig. 1. Tentative genome organization of HCV according to Takamizawa *et al.* (1991) and mapping of cDNA clone. Nucleotides are numbered from the extreme 5' end as indicated at the top. A proposed genetic organization of the HCV genome is indicated, with coding regions for the C, M and E protein as well as nonstructural proteins (NS1 to NS5) shown as block, the 5' as well as 3' non-coding regions as bars. A thin box below HCV genome represent cDNA clones from PCR products which has been amplified from the serum of the HCV infected patients in Nagasaki Prefecture.

Comparison of nucleotide and amino acid sequences: The amino acid sequence was deduced from the nucleotide sequence of the cloned HCV cDNA, and the sequence of the E protein region was tentatively assigned by analysis of the putative cleavage sites and the hydrophobicity profiles of the encoded proteins according to the method reported by Takamizawa *et al.* (1991). The nucleotide and amino acid sequences of the E protein region of HCV cDNAs isolated in Nagasaki Prefecture and others were compared. These comparisons were carried out with computer analysis by using DNASIS-Mac Version 1.0 software system (Hitachi Co., 1991).

RESULTS

Isolation of the cDNA clones of HCV: by RT-PCR, target DNA fragments could be amplified from five patients' serum samples and were isolated as five distinct cDNA clones of HCV in Nagasaki Prefecture, designated as HCV-JN1～HCV-JN5. Brief clinical profile of each patient is showed in Table 1. Two of them were acute hepatitis with history of blood-borne infection (HCV-JN1 and HCV-JN5). Remaining three were chronic hepatic diseases without apparent history of blood transfusion, therefore their routs of infection were unknown.

Nucleotide sequences of the E gene of isolated HCV cDNAs: Figure 2 showed the nucleotide sequence of the E gene of HCV-JN1, an isolate in Nagasaki Prefecture, as well as the nucleotides that differ from this sequence for the four isolates in Nagasaki Prefecture (HCV-JN2～JN5), the five isolates in other parts of Japan; HCV-JYK (Takeuchi *et al.*,

Table 1. Clinical profiles of the patients from whose sera HCVs in Nagasaki Prefecture have been cloned by PCR

Pt No. (virus isolate)	Age/Sex	Date of serum sampling	Clinical diagnosis	History of blood transfusion	ALT level at sampling	Anti C−100 reactivity
1 (HCV−JN1)	49/male	'86. 7. 9	PT−AH*a	(+)	1,925	(−)
2 (HCV−JN2)	58/male	'87.11.19	CH	(−)	114	(++)
3 (HCV−JN3)	53/male	'87.10.22	HCC+LH	(−)	97	(++)
4 (HCV−JN4)	55/male	'87. 7.15	CH	(−)	117	(++)
5 (HCV−JN5)	26/male	'86.12. 6	AH*b	(−)	1,794	(−)

The abbreviations used in clinical diagnosis represents AH; acute hepatitis, PTAH; post-transfusion acute hepatitis, CH; chronic hepatitis, HCC; hepatocellular carcinoma, LC; liver cirrhosis respectively.

*a: HCV has been transmitted from the blood transfused during gingival cancer operation.

*b: This patient is a clinical doctor and HCV has been transmitted from a post-transfusion acute hepatitis patient through the syringe needle.

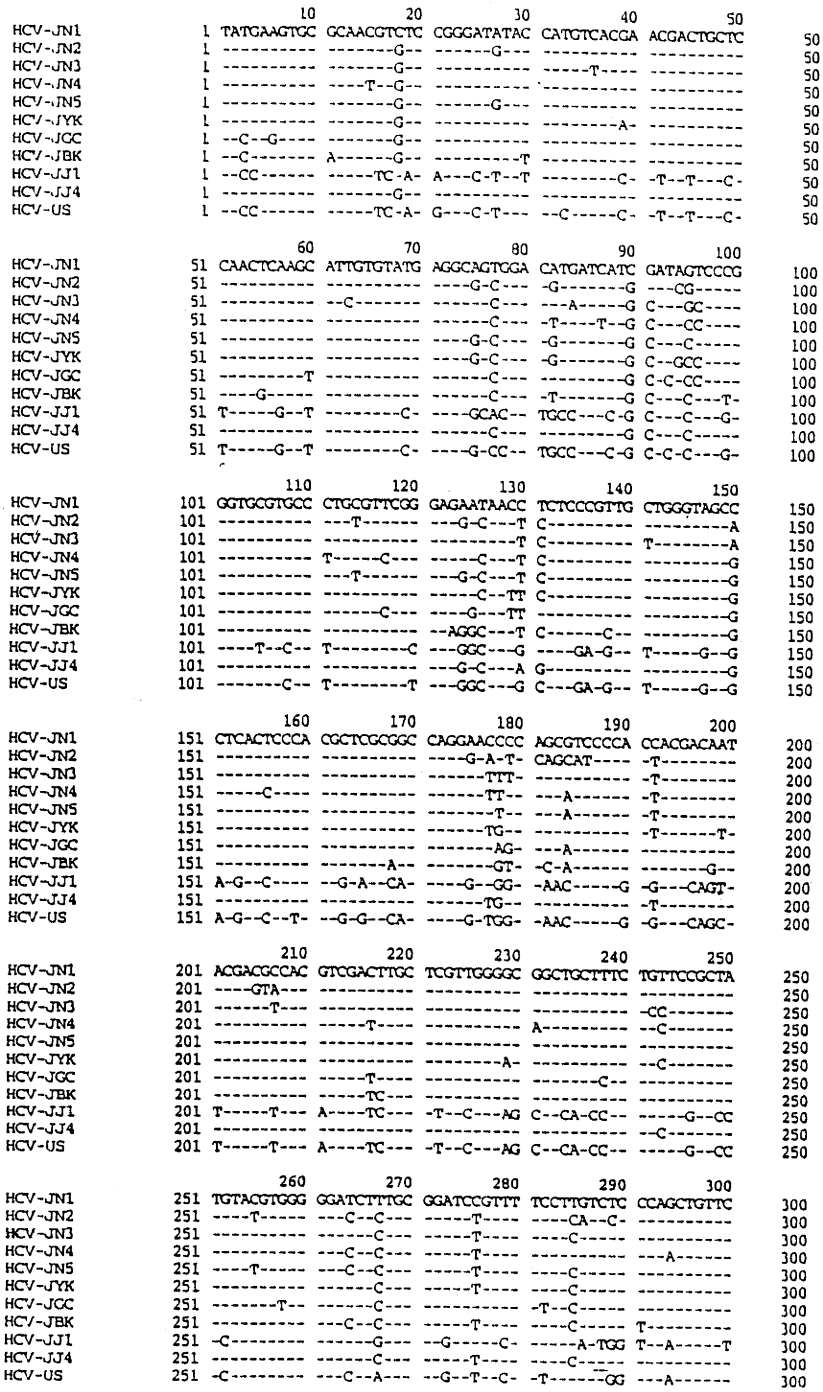

Fig. 2. Nucleotide sequences of the E genes of eleven HCV isolates, HCV-JN1~JN5, HCV-JYK, HCV-JGC, HCV-JBK, HCV-JJ1, JJ2, and HCV-US (nucleotides were numbered from the extreme 5' end of the E gene). Only the changes from HCV-JN1 sequences are indicated for the other isolates. The identical nucleotides are given as bars.

```
                          310        320        330        340        350
HCV-JN1     301 ACCTTCTCGC CTCGCCCGTA TGAGACAGTG CAGGACTGCA ATTGCTCACT 350
HCV-JN2     301 --------A- -G------C- ---------A ---------- -------GA- 350
HCV-JN3     301 ---------- --------C- C-----G--A --------T- --------A- 350
HCV-JN4     301 ---------- ----T---C- C-----GT-- ---------- ----T---A- 350
HCV-JN5     301 --------A- -G-A----C- ---------A ---------- -------GA- 350
HCV-JYK     301 ---------- --------C- ---------A ---------- -C------A- 350
HCV-JGC     301 --------A- ---------- ------G--A --A--T---- --------A- 350
HCV-JBK     301 ---------- --------C- --T----T-A --------T- -C------A- 350
HCV-JJ1     301 --------T- -CA-G--CC- CTG----AC- --A-G----- ----T--TA- 350
HCV-JJ4     301 ---------- --------C- ---------- ---------- -C------A- 350
HCV-US      301 --------T- -CA-G--CC- CTG---GAC- --A-GT---- -------TA- 350

                          360        370        380        390        400
HCV-JN1     351 CTATCCCGGC CATGTAACAG GTCATCGCAT GGCTTGGGAT ATGATGATGA 400
HCV-JN2     351 ---------- T-CC--T--- -------T-- ---------- ---------- 400
HCV-JN3     351 ------T--- --CC------ -C--C----- ---------- ---------- 400
HCV-JN4     351 ---------- --C--GT--- ----G----- ---------C ---------- 400
HCV-JN5     351 ---------- T-CC--T--- -------T-- ---------- ---------- 400
HCV-JYK     351 ---------- --C---T--- -C-------- ---------- ---------- 400
HCV-JGC     351 ---------- --C---T--- ----C----- ---------- ---------- 400
HCV-JBK     351 ---------- -----GT-G- ----C--T-- ---------C ---------- 400
HCV-JJ1     351 ---C------ ---A----G- ---------- ---A------ ---------- 400
HCV-JJ4     351 ---------- ---T--T--- ----C----- ---------- ---------- 400
HCV-US      351 ---------- ---A----G- ----C----- ---A------ ---------- 400

                          410        420        430        440        450
HCV-JN1     401 ACTGGTCACC TACAGCAGCG CTAGTGGTGT CGCAATTACT CCGGATCCCA 450
HCV-JN2     401 ---------- C--------C --G------- ----G--G-- ---------- 450
HCV-JN3     401 -------G-- -------A-C ---------- ----G----- ---------- 450
HCV-JN4     401 ---------- ---TA----C ---------- ----G----- ---------- 450
HCV-JN5     401 ---------- C--------C --G------- ----G--G-- ---------- 450
HCV-JYK     401 -------G-- C--G-----C T--------- ----G----- ---------- 450
HCV-JGC     401 ---------- ----A-G--C --------A- ----G----- ---------- 450
HCV-JBK     401 -------G-- C---A----C ---------- ----G----- ---------- 450
HCV-JJ1     401 -------C-- ---G--G--- T-G--AA--G -T--GC-G-- ---------- 450
HCV-JJ4     401 ---------- ----A----C ---------- ----G--G-- ---------- 450
HCV-US      401 -------C-- ---GA-G--- T-G--AA--G -T--GC-G-- ---------- 450

                          460        470        480        490        500
HCV-JN1     451 CAAGCCGTCG TGGACATGGT GGCTGGGGGC CACTGGGGGG TTCTGGCGGG 500
HCV-JN2     451 -----T---A ---------- ---G--A-C- --------A- -C-------- 500
HCV-JN3     451 ------A--A ---------- -A-G----C- --------A- -C------CC 500
HCV-JN4     451 -----T---- ----T----- -A-G----C- --------AA -C-------- 500
HCV-JN5     451 -----T---A ---------- ---G--A-C- --------A- -C-------- 500
HCV-JYK     451 -----T---A ---------- ---T----C- --------A- -C--A----- 500
HCV-JGC     451 ---------- ---------- ---G----C- --------T- -C--A----- 500
HCV-JBK     451 ---------- ---------- ---G----C- --------A- -C-------- 500
HCV-JJ1     451 ------A--T ----T---A- C-----T-CT --------A- -C-------- 500
HCV-JJ4     451 -----T---- ---------- ---G----C- --------A- -C-------- 500
HCV-US      451 ------A--T --------A- C-----T-CT --------A- -C-------- 500

                          510        520        530        540        550
HCV-JN1     501 GCTCGCCTAC TATTCCATGG TGGGGAACTG GGCTAAGGTC TTGATTGTGA 550
HCV-JN2     501 C--T------ ---------- C--------- ---------T ---------- 550
HCV-JN3     501 CT-G------ ---------- ----A----- ---------- ---------- 550
HCV-JN4     501 C--T--T--- ---------- ---------- ---C-----T --A------- 550
HCV-JN5     501 C--T------ ---------- ---------- ---------T ---------- 550
HCV-JYK     501 C--T------ ---------- ---------- ---------T ---------- 550
HCV-JGC     501 C--T------ ---------- ---------- ---------- ---------- 550
HCV-JBK     501 C--T------ ---------- C--------- ---------T C--------- 550
HCV-JJ1     501 CA-A--G--T -TC------- ---------- ---G------ C--G-A---C 550
HCV-JJ4     501 C--T------ ---------- -A-------- ---------- C--------G 550
HCV-US      501 CA-A--G--T -TC------- ---------- ---G------ C--G-A---C 550

                          560        570        580        590        600
HCV-JN1     551 TGTTACTCTT TGCCGGCGTT GACGGAAGCA CCGTCGTGAC AGGG...... 600
HCV-JN2     551 --C------- C--------- --T--GCAA- --CG------ G--------- 600
HCV-JN3     551 --C-G--T-- ---------- -----GGAT- --CG------ ---------- 600
HCV-JN4     551 --C----T-- ---------- -----GG-A- --CA----T- ---------- 600
HCV-JN5     551 --C------- C--------- --T-CGGCA- A-CG------ G--------- 600
HCV-JYK     551 --C------- ---------- -----GCAT- --CG------ G--------- 600
HCV-JGC     551 --C------- ---T------ -----GCA-- --CA------ ---------- 600
HCV-JBK     551 --C----T-- ---T------ -----GGAT- --CA------ ---------- 600
HCV-JJ1     551 ----G--G-- ---------C ----CGGAA- --A----CT- C--------- 600
HCV-JJ4     551 C-C------- ---------- -----GGAG- --TA-AC-T- G--------- 600
HCV-US      551 --C-G--A-- ---------C ----CGGAA- --CA---C-- C--------- 600
```

1990), HCV-JGC (Kato *et al.*, 1990), HCV-JBK (Takamizawa *et al.*, 1991), HCV-JJ1 and HCV-JJ4 (Okamoto *et al.*, 1990), and the prototype isolate in the United States; HCV-US (Choo *et al.*, 1991). This figure showed that the nucleotide differences were scattered throughout the length of the gene, and there appeared to be no particular region of hypervariability. The numbers of nucleotide differences and the identity percent of the necleotidies between each of these isolates were summarized in Table 2 (a) and (b), respectively. These results indicated that the isolate in the United States (HCV-US) was less closely related to the isolates in Japan except for HCV-JJ1. Except HCV-JJ1, HCV-US showed an average of 25.5% nucleotide sequence difference with all Japanese HCV cDNA islates, and the difference among Japanese isolates was in the range from 4.2% to 12.1% (average 9.1%). HCV-JJ1, which was isolated from a HCV healthy carrier in Japan, was closely related to HCV-US showing 94.6% nucleotide sequence identity. Although pairs of other isolates in Japan showed similar nucleotide sequence identity each other, there was a pair of isolates which showed less nucleotide difference; HCV-JN2 and HCV-JN5, both of which were isolated in Nagasaki Prefecture, showed a 96.3% identity.

Amino acid differences in the E protein of HCV isolates: The encoded amino acid sequences of the E protein from HCV-JN1 isolate was shown, in the standard one letter code, in Figure 3, with different amino acids noted for the other isolates. Five potential N-glycosylation sites were found from the amino acid sequence of the E protein of HCV-JN1. Two of them were completely conserved in all other isolates. In the remaining three sites, amino acid sequences were different in some other isolates. But in two of these sites, the pattern of the amino acid sequence (N-X-T/S) were conserved in all other isolates. Various amino acid differences among HCV strains were scattered throughout the E protein, and there appeared to be no particular region of hypervariability. Table 2 (a) and (b) summarized the amino acid differences and identities percent between these isolates. These results indicated that HCV-US and HCV-JJ1 were less closely related to other Japanese isolates, similar to the nucleotide sequences comparison. The amino acid identity percent between each isolate were similar to those of the nucleotide identity indicating that most of the mutations were not silent. The difference between the amino acid sequence homology percent calculated by using the Dayhoff scoring matrix (Dayhoff and Schwartz, 1978) and the amino acid sequences identity percent were shown in Table 3. Homology among each isolates including also HCV-US and HCV-JJ1 showed high matching figure (average 95.9%), compared with those of the identity (average 87.1%) indicating that many of the mutations were not silent, but were neutral (conservative). Figure 4 showed the amino acid sequence of the E protein of HCV-JN1 isolate, as well as the amino acids that were not homologous with this sequence for the other HCV isolates. There appeared to be three conserved regions (AA no.1～25, no.81～106, no.122～192). Figure 5 showed the relation of the hydrophobicity profile of the E protein of HCV-JN1 clone using Hopp and Woods's program (1981) that is useful for prediction of protein antigenic determinants, and the locations of these three conserved regions. Several hydrophilic portions were observed on region A, but few on region B and C.

Table 2. (a) Nucleotide and predicted amino acid differences among eleven HCV isolates in the gene coding for the envelope protein

NUCLEOTIDE \ AMINO ACID	HCV−JN1	HCV−JN2	HCV−JN3	HCV−JN4	HCV−JN5	HCV−JYK	HCV−JGC	HCV−JBK	HCV−JJ1	HCV−JJ4	HCV−US
HCV−JN1	−	25	20	20	22	17	14	22	49	16	47
HCV−JN2	71	−	19	21	12	15	19	20	47	16	46
HCV−JN3	60	65	−	17	16	14	17	20	47	14	44
HCV−JN4	65	72	57	−	17	18	13	13	48	13	43
HCV−JN5	61	25	58	56	−	14	14	19	47	12	41
HCV−JYK	55	48	44	52	39	−	13	18	44	12	43
HCV−JGC	52	71	57	57	56	49	−	14	46	11	41
HCV−JBK	68	67	62	52	56	54	54	−	47	16	43
HCV−JJ1	157	167	166	160	158	158	160	156	−	44	5
HCV−JJ4	49	58	51	48	44	37	50	49	156	−	43
HCV−US	160	158	153	151	151	150	144	147	33	147	−

Numbers above the diagonal represent amino acid differences among the HCV isolates, while those below the diagonal represent nucleotide differences.

(b) Nucleotide and predicted amino acid identity % among eleven HCV isolates in the gene coding for the envelope protein

NUCLEOTIDE \ AMINO ACID	HCV−JN1	HCV−JN2	HCV−JN3	HCV−JN4	HCV−JN5	HCV−JYK	HCV−JGC	HCV−JBK	HCV−JJ1	HCV−JJ4	HCV−US
HCV−JN1		93.9%	96.0%	97.0%	96.5%	97.0%	96.5%	95.0%	93.4%	96.5%	92.%4
HCV−JN2	87.9%		96.5%	96.0%	96.5%	96.5%	94.4%	97.0%	91.4%	96.0%	93.0%
HCV−JN3	90.0%	89.4%		98.0%	97.5%	98.5%	96.5%	97.5%	92.9%	98.0%	94.4%
HCV−JN4	89.1%	88.4%	90.6%		98.5%	99.0%	97.0%	97.0%	93.0%	98.0%	94.4%
HCV−JN5	89.6%	96.3%	90.2%	90.4%		98.5%	96.5%	96.5%	92.4%	97.5%	93.9%
HCV−JYK	90.6%	92.4%	92.6%	91.4%	93.6%		98.0%	97.5%	93.4%	98.5%	95.0%
HCV−JGC	91.1%	88.6%	90.2%	90.2%	89.9%	91.6%		96.5%	94.0%	96.5%	95.0%
HCV−JBK	88.4%	88.9%	89.4%	90.4%	90.2%	90.9%	90.6%		93.4%	97.0%	95.0%
HCV−JJ1	73.6%	72.6%	72.2%	73.1%	73.4%	73.4%	73.1%	73.6%		93.4%	98.5%
HCV−JJ4	91.6%	90.7%	91.4%	91.8%	92.6%	93.6%	91.3%	91.4%	73.6%		94.4%
HCV−US	73.1%	73.7%	74.2%	74.8%	74.4%	74.4%	75.6%	75.1%	94.6%	75.1%	

Numbers above the diagonal represent amino acid identity % among the HCV isolates, while those below the diagonal represent nucleotide identity %.

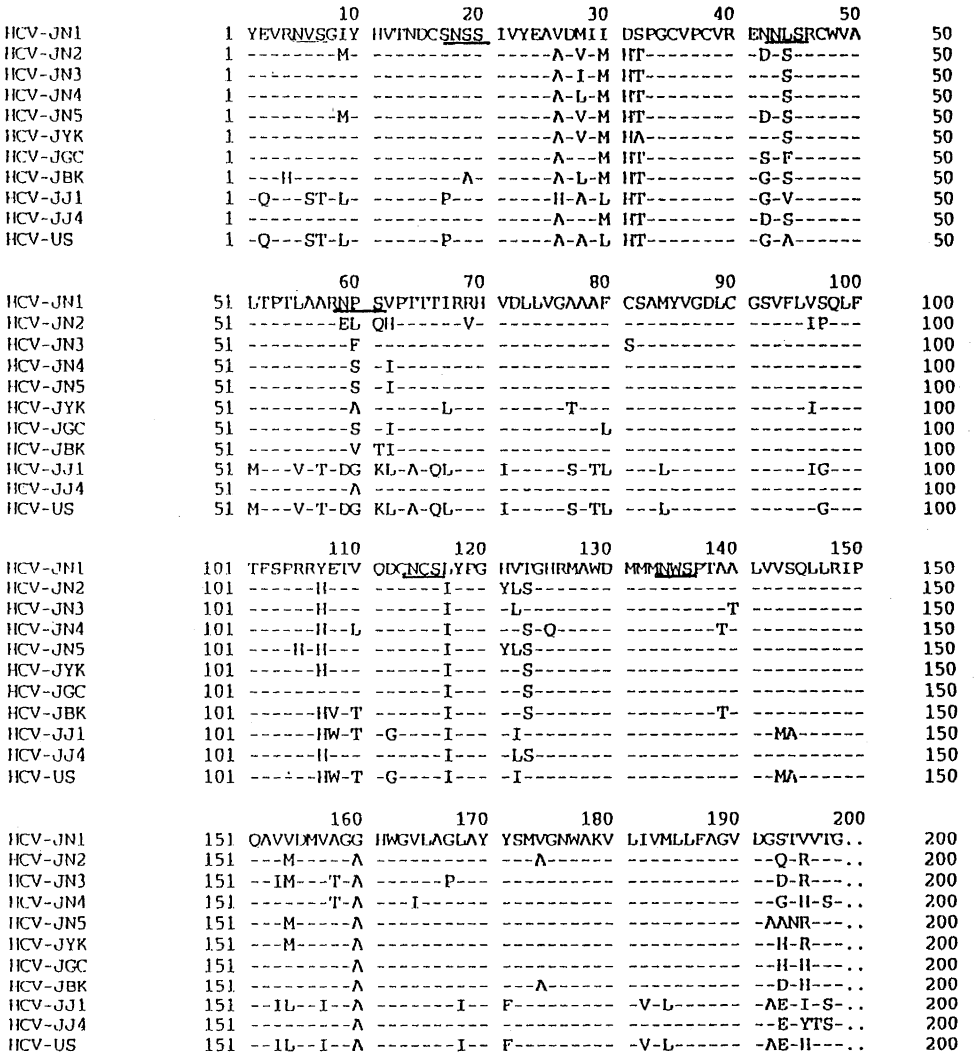

Fig. 3. Envelope protein sequences derived by the translation of the nucleotide sequences shown in Fig. 1 of eleven HCV isolates (amino acid were numbered from the extreme 5' end of the E protein). Only the amino acids that differ from HCV-JN1 are indicated for the other isolates. The identical amino acids are given as bars. The five potential N-glycosylation sites (N-X-T/S) are underlined.

Table 3. Amino acid homology % and identity % among eleven HCV isolates of the envelope protein

AMINO ACID HOMOLOGY

AMINO ACID IDENTITY

	HCV−JN1	HCV−JN2	HCV−JN3	HCV−JN4	HCV−JN5	HCV−JYK	HCV−JGC	HCV−JBK	HCV−JJ1	HCV−JJ4	HCV−US
HCV−JN1		93.9%	96.0%	97.0%	96.5%	97.0%	96.5%	95.0%	93.4%	96.5%	92.%4
HCV−JN2	87.4%		96.5%	96.0%	96.5%	96.5%	94.4%	97.0%	91.4%	96.0%	93.0%
HCV−JN3	89.9%	90.4%		98.0%	97.5%	98.5%	96.5%	97.5%	92.9%	98.0%	94.4%
HCV−JN4	89.9%	89.4%	91.4%		98.5%	99.0%	97.0%	97.0%	93.0%	98.0%	94.4%
HCV−JN5	88.9%	93.9%	92.0%	91.4%		98.5%	96.5%	96.5%	92.4%	97.5%	93.9%
HCV−JYK	91.4%	92.4%	92.9%	91.0%	92.9%		98.0%	97.5%	93.4%	98.5%	95.0%
HCV−JGC	92.9%	90.4%	91.4%	93.4%	92.9%	93.4%		96.5%	94.0%	96.5%	95.0%
HCV−JBK	88.9%	89.9%	89.9%	93.4%	90.4%	91.0%	92.9%		93.4%	97.0%	95.0%
HCV−JJ1	75.3%	76.3%	76.3%	75.8%	76.3%	77.8%	76.8%	76.3%		93.4%	98.5%
HCV−JJ4	92.0%	92.0%	92.9%	93.4%	93.9%	93.9%	94.4%	92.0%	77.8%		94.4%
HCV−US	76.3%	76.8%	77.8%	78.3%	79.3%	78.3%	79.3%	78.3%	97.5%	78.3%	

Numbers above the diagonal represent amino acid homology % among the HCV isolates, while those below the diagonal represent nucleotide identity % same as numbers above the diagonal in Table 2 (b).

Evolutional relationships among the HCV isolates: According to the method to calculate evolutional difference (*Ks*) between two homologous nucleotide sequences reported by Miyata and Yasunaga (1980), *Ks* for each pair of the compared HCV cDNA isolated has been calculated by using "EVO-DNA4" computer program (Okumura and Izui, 1990) (the program soft has been provided by Dr. K. Izui, Department of Chemistry, Faculty of Science, Kyoto University). Based on the *Ks* values, phyrogenetic relatinships among compared HCV cDNA isolates were determined by the unweighted pair-group method (Sneath and Sokal, 1973) (Fig. 6). From this result, HCV cDNA isolates compared in this study were roughly divided into two groups; the first group contains HCV-US and HCV-JJ1 and the second group contains the remaining isolates.

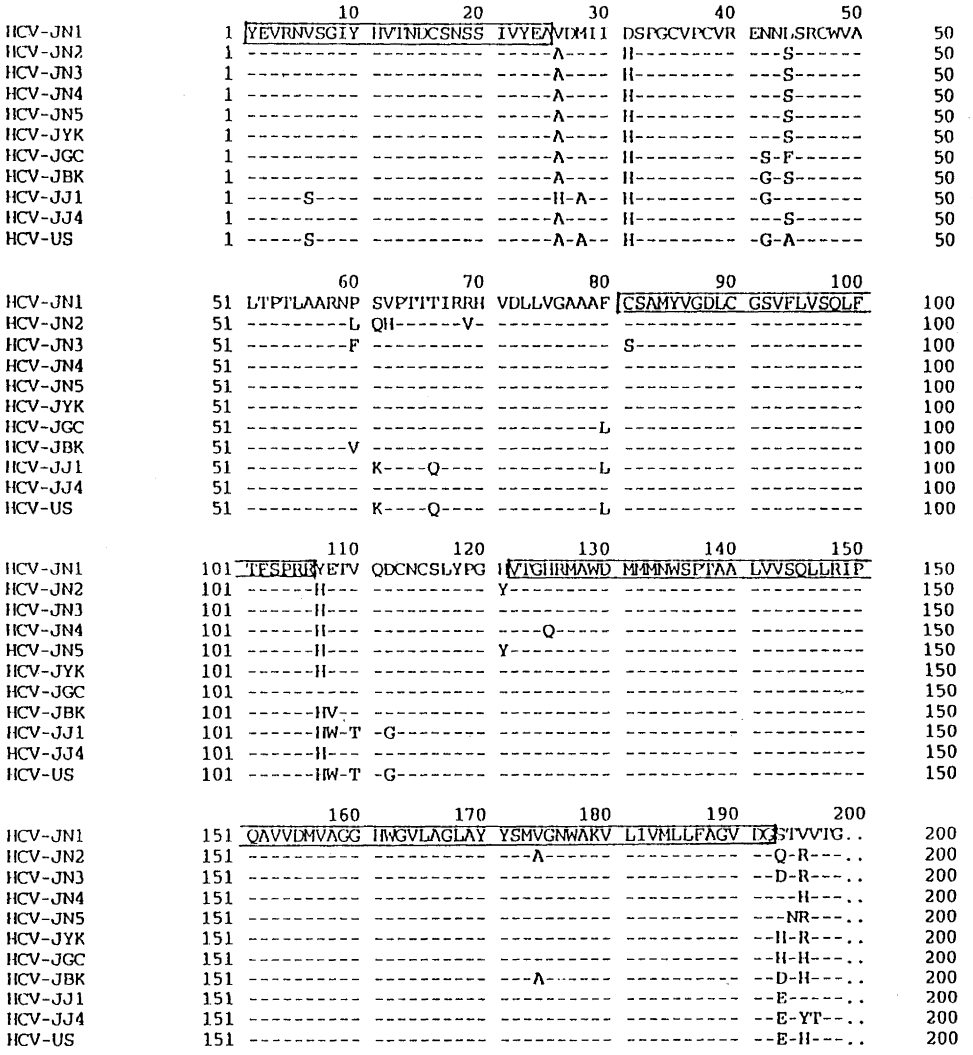

Fig. 4. Envelope protein sequences of eleven HCV isolates as same as shown in Fig. 3. The amino acids that are not homologous to HCV-JN1 are indicated. The homologous amino acids are given as bars. The sequence regions in where amino acids homology is highly conserved among the all eleven isolates are boxed.

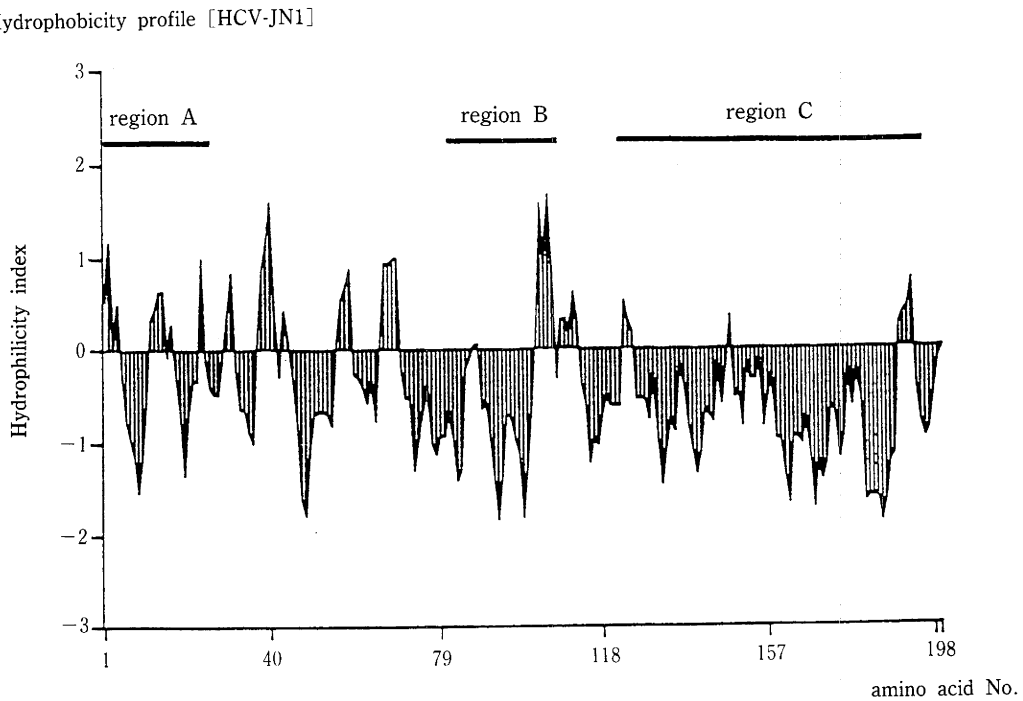

Fig. 5. Hydrophilicity profile of E protein from HCV-JN1. The bold bars given as region A, B, C indicate the location of the highly conserved sequence regions shown as boxed regions in Fig. 4.

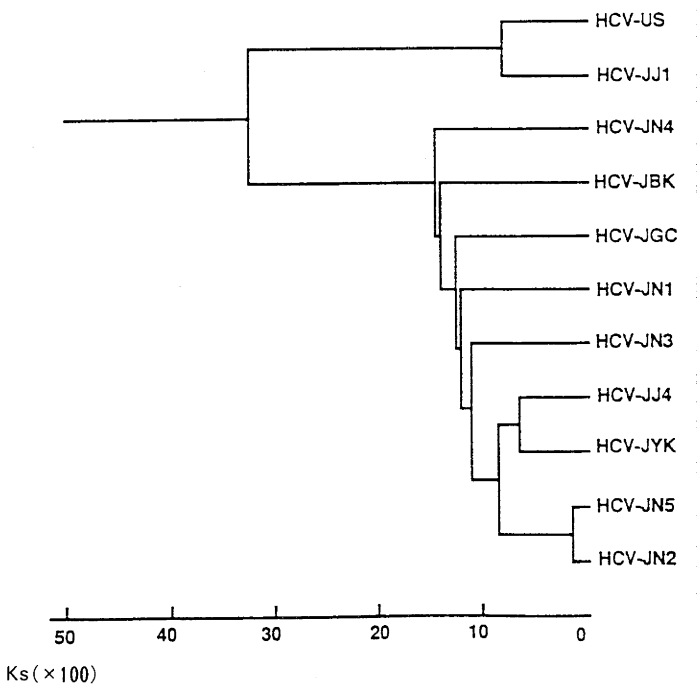

Fig. 6. Phyrogenetic tree illustrating the evolutional relationship of the eleven HCV isolates.

Discussion

The comparison of nucleotide and amino acid sequences among various HCV cDNA isolates and the phyrogenetic tree calculated from these results showed that there was no closer relationships among the isolates in Nagasaki Prefecture compared with those in other parts in Japan. The result indicated that the compared HCV cDNA isolates in Japan, except HCV-JJ1, were included in one large subtype (Japanese subtype), and a isolate in the United States (HCV-US) was distinct from this subtype. HCV-JJ1 was considered to belong to the HCV-US because of its sequence relatedness. The nucleotide and amino acid differences were scattered throughout the length of the E protein region, and no particular hypervariable or conserved region was observed. Although the amino acid identity between each pair of the compared isolates showed some variation (75.3%～97.5%, average 87.1%), the amino acid homology based on the Dayhoff scoring matrix was well conserved (91.4%～99.0%, average 95.9%) among all isolates including US subtype strains. While seve ·l highly conserved region were observed in the E protein sequence, indicating that many of amino acid mutations in the E protein region were conservetive. This result suggests that conformation of the HCV E protein is relatively conserved, indicating that the E protein region may be under structural or functional constraint. From the comparison with hydrophobicity profile of the E protein, one of the conserved region (region A in Fig. 5) may possibly contain the antigenic domain which function as neutralizing epitopes. Recently, Matsuura *et al.* (1991) expressed the HCV E protein gene as a 35 kDa glycoprotein (gp35) by using a recombinant baculovirus. By Western blot analysis using expressed gp35 as antigen, they tried to detect antibodies to gp35 in the sera from NANBH patients. From this experiment, they got following observations. Antibodies to gp35 were detectable in only 4%～11% of NANBH patients, not only in acute hepatitis but also chronic hepatic diseases (chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma). In two of the acute hepatitis and one of the chronic hepatitis cases, patients have been recovered from hepatitis after appearance of anti-gp35 antibodies. The serum alanin aminotransferase level has been generally kept at lower level in the anti-gp35 seropositive chronic disease patients. These results raise two possibilities: (i) gp35 does not contain neutralizing epitopes or has very low immonogenicity, but has some functions to inhibit viral proliferation, (ii) although gp35 contains neutralizing epitopes, their antigenic structure is different among various HCV because of the hypervariability of amino acid sequences. From the results presented in this parer, the second possibility does not seem to be probable. On the other side, some researchers proposed that the hypervariable region which exists in the NS1 region carries potential neutralizing epitopes and suggested the mechanism resembling to the second possibility mentioned above that make lead to the persistent infection of HCV (Hijikata *et al.*, 1991; Kremsdolf *et al.*, 1991; Weiner *et al.*, 1991).

At present, several speculations and suggestions are under discussion concerning this problem. To get more precise and useful information, further experiments on HCV and establishment of the cell culture system for HCV will be necessary.

ACKNOWLEDGEMENTS

The author would like to express thanks to Prof. A. Igarashi for valuable advice and encouragement, careful reading and correcting the manuscripts.

This study was supported by the grants for the NANBH study group of the Ministry of Health and Welfare, Japan.

REFERENCES

1) Alter, H. J., Purcell, R. H., Shih, J. M., Melpolder, J. C., Houghton, M., Choo, Q-L. & Kou, G. (1989): Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. N. Engl. J. Med., 321, 1494-1500.

2) Choo, Q-L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P.J., Weiner, A.J., Bradley, D.W., Kou, G. & Houghton, M. (1991): Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. USA, 88, 2451-2455.

3) Choo, Q-L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. & Houghton, M. (1989): Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science, 244, 359-362.

4) Dayhoff, M.O. & Schwaltz, R.M. (1978): Atlas of protein sequence and structure, (ed.) by Dayhoff, M.O., National Biomedical Research Foundation, Washington, 5, Supple., 3, 353.

5) Dienstag, J.L. (1983): Non-A, non-B hepatitis. I. recognition, epidemiology, and clinical features. Gastroenterol., 85, 439-462.

6) Feinstone, S.M., Kapikian, A.Z., Purcell, R.H., Alter, H.J. & Holland, P.V. (1975): Transfusion-associated hepatitis not due to viral hepatitis type A or B. N. Engl. J. Med., 292, 767-770.

7) Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. & Shimotohno, K. (1991): Hypervariable region in the putative glycoprotein of hepatitis C virus. Biochem. Biophys. Commun., 175, 220-228.

8) Hopp, T.P. & Woods, K.R., (1981): Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA, 78, 3824−3828.

9) Houghton, M., Choo, Q-L. & Kuo, G. (1988): Eur. Patent Appl. 88, 310, 922.5 and Publ. 0, 318, 216.

10) Kato, N., hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990): Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. USA, 87, 9524-9528.

11) Kiyosawa, K., Akahane, Y., Nagata, A. & Furuta, S. (1984): Hepatocellular carcinoma after non-A, non-B posttransfusion hepatitis. Am. J. Gastroenterol. 79, 777-781.

12) Koretz, R.L., Stone, O. & Gitnick, G.L. (1980): The long-term course of non-A, non-B post-transfusion hepatitis. Gastroenterol., 79, 893-898.

13) Kremsdorf, D., Porchon, C., Kim, J.P, Reyes, G.R. & Brechot, C. (1991): Partial nucleotide sequence analysis of a French hepatitis C virus: implication for HCV genetic variability in the E2/NS1 protein. J. Gen. Virol., 72, 2557-2561.

14) Kuo, G., Choo, Q-L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., Tegtmeieer, G.E., Bonino, F., Colombo, M., Lee, W-S., Kuo, C., Berger, K., Shuster, J.R., Overby, L.R., Bradley, D.W. & Houghton, M. (1989): An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. Science, 244, 362-364.

15) Matsuura, Y., Shizuko, H.,Watanabe, Y., Chiba, T., Saito, I. & Miyamura, T. (1991): Expression of hepatitis C virus protein. Saibou Kogaku, 10, 845-852.

16) Miller, R. H. & Purcell, R. H. (1990): Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus super groups. Proc. Natl. Acad. Sci. USA, 87, 2057-2061.

17) Miyamura, T., Saito, I., Katayama, T., Kikuchi, S., Tateda, A., Houghton, M., Choo, Q-L. & Kuo, G. (1990): Detection of antibody against antigen expressed by molecularly cloned hepatitis C virus cDNA: Application to diagnosis and blood screening for posttransfusion hepatitis. Proc. natl. Acad. Sci. USA, 87, 983-987.

18) Miyata, T., Hayashida, H., Kikuno, R. & Yasunaga, T. (1986): The method to calculate evolutional difference between two homologous nucleotide sequences pp. 404-425. In The Japanese Biochemical Society (ed.). The Sequel Lecture of Biochemical Experiment. Tokyo Kagaku-Dojin. Tokyo.

19) Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1990): The 5'-terminal sequence of the hepatitis C virus genome. Japan. J. Exp. Med., 60, 167-177.

20) Okumura, S. & Izui, K. (1990): A BASIC program to calculate evolutional difference between two homologous nucleotide sequences. Seikagaku, 62, 1385-1389.

21) Plagemann, P.G.W. (1991): Hepatitis C. Brief review. Arch. Virol., 120, 165-180.

22) Realdi, G., Alberti, A., Rugge, M., Rigoli, A.M., Tremolada, F., Schivazappa, L. & Roul, A. (1982): Long-term follow-up of acute and chronic non-A, non-B post-tramsfision hepatitis: evidence of progression to liver cirrhosis. Gut, 23, 270-275.

23) Resnick, R.H., Stone, K. & Antonioli, D. (1983): Primary hepatocellular carcinoma following non-A, non-B posttransfusion hepatitis. Dig. Dis. Sci., 28, 908-911.

24) Saiki, R.K., Gelefand, D.H., Stofeel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239, 487-491.

25) Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, TY., Koi, S., Onji, M., Ohta, Y., Choo, Q-L., HOughton, M. & Kuo, G. (1990): Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. Proc. Natl. Acad. Sci. USA, 87, 6547-6549.

26) Sanger, P., Nicklen, S & Coulson, A.R. (1977): DNA sequencing with chain-termination inhibitors. Proc. Natl. Acad. Sci. USA, 74, 5463-5467.

27) Sneath, P.H.A. & Sokal, R.R. (1973): Principles of numerical Taxonomy. W.H. Freeman, San Francisco.

28) Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E, Andoh, T., Yoshida, I. & Okayama, H. (1991): Structure and organization of the hepatitis C virus genome isolated from human carriers. J. Virol., 65, 1105-1113.

29) Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990): The putative nucleocapsid and envelope protein genes of hepatitis C virus determined by comparison of the nucleotide sequences of two isolates derived from an experimentally infected chimpanzee and healthy human carriers. J. Gen. Virol., 71, 3027-3033.

30) Alter, A.J., Brauer, M.T., Rosenblatt, J., Richman, K.H., Tung, J., Crowford, K., Bonino, F., Saracco, G., Choo, Q-L., Houghton, M. & Han, J.H. (1991): Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology, 180, 842-848.